STEM CELLS

STEM CELL TECHNOLOGY: EPIGENETICS, GENOMICS, PROTEOMICS AND METABONOMICS

Dynamic Proteomic Profiling of Extra-Embryonic Endoderm Differentiation in Mouse Embryonic Stem Cells

CLAIRE M. MULVEY,[a,b,c] CHRISTIAN SCHRÖTER,[c] LAURENT GATTO,[a,b,d] DUYGU DIKICIOGLU,[b,e] ISIK BARIS FIDANER,[f] ANDY CHRISTOFOROU,[a,b,e] MICHAEL J. DEERY,[a,b] LILY T. Y. CHO,[g] KATHY K. NIAKAN,[h] ALFONSO MARTINEZ-ARIAS,[c] KATHRYN S. LILLEY[a,b,e]

Key Words. Differentiation • Extra-embryonic endoderm stem cells • Pluripotency • Quantitative proteomics • Tandem mass tags

[a]Cambridge Centre for Proteomics, Department of Biochemistry, [b]Cambridge Systems Biology Centre, Wellcome Trust Stem Cell building, [c]Department of Genetics, [d]Computational Proteomics Unit, Department of Biochemistry, [e]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom; [f]Department of Computer Engineering, Bogazici University, Istanbul, Turkey; [g]Neusentis, Pfizer Worldwide Research and Development, Granta Park Science Park, Great Abington, Cambridge, United Kingdom; [h]The Francis Crick Institute, Mill Hill Laboratory, London, United Kingdom

Correspondence: Alfonso Martinez-Arias, Ph.D., Department of Genetics, University of Cambridge, Cambridge CB2 1QR, U.K. Telephone: +44 (0)1223 766742; e-mail: ama11@hermes.cam.ac.uk; or Kathryn S. Lilley, Ph.D., Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR, U.K. Telephone: 01223 760255; e-mail: k.s.lilley@bioc.cam.ac.uk

Received December 17, 2014; accepted for publication April 20, 2015; first published online in STEM CELLS *EXPRESS* June 23, 2015.

1066-5099/2015/$30.00/0

http://dx.doi.org/10.1002/stem.2067

ABSTRACT

During mammalian preimplantation development, the cells of the blastocyst's inner cell mass differentiate into the epiblast and primitive endoderm lineages, which give rise to the fetus and extra-embryonic tissues, respectively. Extra-embryonic endoderm (XEN) differentiation can be modeled in vitro by induced expression of GATA transcription factors in mouse embryonic stem cells. Here, we use this GATA-inducible system to quantitatively monitor the dynamics of global proteomic changes during the early stages of this differentiation event and also investigate the fully differentiated phenotype, as represented by embryo-derived XEN cells. Using mass spectrometry-based quantitative proteomic profiling with multivariate data analysis tools, we reproducibly quantified 2,336 proteins across three biological replicates and have identified clusters of proteins characterized by distinct, dynamic temporal abundance profiles. We first used this approach to highlight novel marker candidates of the pluripotent state and XEN differentiation. Through functional annotation enrichment analysis, we have shown that the downregulation of chromatin-modifying enzymes, the reorganization of membrane trafficking machinery, and the breakdown of cell–cell adhesion are successive steps of the extra-embryonic differentiation process. Thus, applying a range of sophisticated clustering approaches to a time-resolved proteomic dataset has allowed the elucidation of complex biological processes which characterize stem cell differentiation and could establish a general paradigm for the investigation of these processes. STEM CELLS *2015;33:2712—2725*

SIGNIFICANCE STATEMENT

As cells specialize to carry out specific tasks during the development of embryos, the collection of proteins that they express (the proteome) changes in a defined and coordinated manner. In this work we map these changes during the targeted differentiation of embryonic stem cells into one of the first specialized cell types that appears in mammalian development. This roadmap of proteome change will be an important reference to better understand how changes in the building blocks of cells drive the change of a cell's function, and will be useful to validate future efforts in which immature stem cells are being differentiated into cell types for regenerative approaches.

INTRODUCTION

The specification of different cell types during development is a complex process that is organized in a hierarchical fashion. Lineage-specific transcription factors activate the expression of batteries of genes which endow cells with particular phenotypes that allow them to change their morphologies and exert defined functions. One of the earliest examples of this process is found in the mammalian preimplantation embryo. At embryonic day 3 (E3.0), the mouse embryo is comprised of an outer epithelial layer, the trophectoderm, that will give rise to the placenta and an inner cell mass (ICM) which will be subdivided into the primitive endoderm (PrE) and the epiblast (Epi). Some extra-embryonic tissues are derived from the PrE, while the Epi develops into the embryo proper. Initially, all cells in the ICM coexpress transcription factors for the Epi fate, such as OCT4, SOX2, and NANOG, together with the transcriptional regulator GATA6 that promotes PrE specification and is a marker for this fate at later stages. Cells within

the ICM that undergo PrE specification progressively downregulate the Epi-specific transcription factors concomitantly with an increase in the expression of GATA6 and other PrE-specific transcriptional regulators such as GATA4 and SOX17 [1–3]. Importantly, in addition to these differences in transcription factor expression, the PrE and the Epi are also morphologically distinct: while the Epi before implantation has mesenchymal character, the PrE initially forms an epithelium lining the Epi. Subsequently, the PrE further differentiates into the epithelial visceral endoderm, a tissue surrounding the post-implantation Epi, and the parietal endoderm, which deposits the extracellular matrix of Reichert's membrane [4, 5].

Some of the early fate decisions of the mammalian embryo can be recapitulated in culture and thus provide a good experimental system to further define and investigate both the transcriptional control as well as the execution of cell fate change. Pre-implantation blastocysts can be used to obtain self-renewing cell populations representing the Epi and the PrE. Embryonic stem cells (ESCs), for instance, share many functional and molecular characteristics of the Epi lineage, and extra-embryonic endoderm (XEN) cells display a molecular signature that resembles the PrE [6, 7]. ESCs and XEN cells have very different morphological characteristics: while ESCs grow in tight colonies, XEN cells have a flat mesenchymal appearance and are highly motile. ESCs grown under specific experimental conditions can contribute to extra-embryonic tissues in chimaera experiments [3, 8, 9], and spontaneous differentiation of XEN-like cells in ESC cultures has been observed [10]. Most significantly, induced expression of GATA factors in ESCs will produce XEN-like cells efficiently [11, 12].

While the role of transcriptional regulators that drive the divergence of pluripotent Epi and PrE specification has been intensively investigated in recent years, the temporal hierarchy of the protein expression programs that they direct and the global changes that result when cells undergo differentiation are only beginning to be elucidated. The most straightforward way of analyzing these global changes is by transcriptomic microarray analysis or RNA-seq (see ref. [10] for an example in the context of differentiation of XEN-like cells). However, changes at the mRNA level do not necessarily reflect changes at the protein expression level [13], emphasizing the need to investigate the execution of cell differentiation with proteomic techniques. Several large-scale proteomics studies have begun to address the processes involved in regulation and maintenance of pluripotency, complementing the wealth of information generated from transcriptomic and epigenetics studies (see ref. [14] for a recent review). However, these studies often use either undirected differentiation protocols, focus on a specific subcellular region, or compare only a few terminally differentiated states [15–17]. We reasoned that the experimentally controlled differentiation of extra-embryonic cells from ESCs through inducible expression of GATA factors would represent an ideal case to delineate global molecular changes in the differentiating proteome and define the dynamic, temporal sequence of events that underlie the early stages of this process.

Using a tandem mass tags (TMT)-differential labeling mass spectrometry-based approach, we have robustly quantified over 2,000 proteins associated with the transition from an ESC- to a XEN-like phenotype following inducible GATA4 expression. A range of multivariate data analysis methods were used to define clusters of proteins whose temporal abundance profiles correlate with morphological changes. This approach has demonstrated the equivalence of GATA4 and GATA6 to implement the ESC-to-XEN transition, and identified novel indicators of both the self-renewing, pluripotent ESC population and the fully differentiated XEN state. We find that the onset of differentiation induced by GATA4 triggers alterations in chromatin remodeling factors, enhanced membrane trafficking, reorganization of extracellular matrix and components of the adherens junctions, as well as altered metabolic capacity. We also show that some effector molecules of ERK/MAPK signaling may be involved in the transition to a more migratory XEN-like phenotype. We anticipate that the results presented here will provide a valuable resource for the further investigation of the cellular-wide changes induced during lineage commitment in mouse ESCs and the early cell fate decision in vivo.

Materials and Methods

Cell Culture, GATA4, and GATA6 Overexpression

ESCs and XEN cells were cultured on gelatin-coated plastic in glasgow minimal essential medium (GMEM) supplemented with 10% ESC qualified fetal bovine serum, nonessential amino acids, sodium pyruvate, stable glutamine, β-mercaptoethanol, and leukemia inhibitory factor. Medium was changed daily, and cells passaged as required upon reaching confluence.

The full-length mouse cDNA for *Gata4* or *Gata6* was amplified from XEN cells, FLAG-tagged and cloned into a plasmid carrying a bidirectional tetracycline-inducible minimal cytomegalovirus (CMV) promoter that simultaneously drives the expression of a gene coding for a red fluorescent protein (dsRed). Coelectroporation in mouse ESCs of this plasmid together with a plasmid expressing the FLPe recombinase allowed for successful targeting to a modified *ColA1* locus, as described previously [18]. The ESCs also contain a modified *ROSA26*-locus that drives doxycycline-inducible expression of an M2rtTA transactivator. Successfully targeted cells were selected using puromycine (1 μg/ml) and hygromycine (50-100 μg/ml), and will be described in further detail elsewhere (Wamaitha S.E. et al., accepted). GATA4-FLAG and GATA6-FLAG expression was induced by addition of 500 ng/ml doxycycline to the culture medium.

Western Blotting and Immunohistochemistry

Cells were harvested in 8 M urea, 0.1% SDS, 25 mM triethyl ammonium bicarbonate (TEAB), pH 8.5, with protease inhibitors (Roche, Burgess Hill, West Sussex, UK, www.roche.co.uk/) and protein concentration was evaluated with the BCA protein assay (Thermo Fisher, Scientific, Loughborough, Leics., UK, www.lifetechnologies.com). To further ensure comparable protein levels between samples, for each of the biological replicates a control blot was probed with an antibody directed against α-tubulin. Between 8 and 20 μg of each sample was loaded onto 4%–15% pre-cast SDS polyacrylamide gel electrophoresis gels (Bio-Rad, Hercules, CA, www.bio-rad.com) and transferred onto nitrocellulose membranes using the Trans Blot Turbo transfer system (Bio-Rad, Hercules, CA, www.bio-rad.com). Primary antibody information can be found in Supporting Information Table S1. Secondary antibodies were fluorescently labeled (IRDye, LI-COR) or horse radish peroxidase (HRP)-conjugated (Goat anti-Rabbit IgG-HRP, BioRad; Donkey

anti-Goat IgG-HRP, Santa Cruz, Dallas, TX, www.scbt.com/); Goat anti-Mouse IgG-HRP, Bio-Rad), and visualized by infrared fluorescence detection (LI-COR, Lincoln, NE, www.licor.com/) or enhanced chemiluminescence (GE Healthcare, Amersham, Bucks, UK, www3.gehealthcare.co.uk/). For reprobing of membranes, antibodies were removed by incubation of the membranes in 2% SDS, 100 mM β-mercaptoethanol, and 20 mM Tris-Cl pH 6.8 for 20 minutes at 65°C.

For immunocytochemistry, cells were plated on ibidi µ-slides, and staining was performed according to standard procedures. Briefly, cells were washed, fixed for 15 minutes in buffered 4% formaldehyde and permeabilized with 0.1% Triton X-100 in the presence of 1% bovine serum albumin. Primary antibodies were diluted in permeabilization/blocking buffer as described in Supporting Information Table S1 and incubated overnight. Secondary antibodies were Alexa-Fluor conjugated IgG from Molecular Probes Loughborough, Leics., UK, www.lifetechnologies.com, and diluted 1:500. Hoechst 33342 was added at 10 µg/ml to the secondary antibody solution to visualize nuclei. Images were captured on a Zeiss LSM700 confocal microscope using a 40X/1.4 NA oil immersion lens.

TMT Labeling, Peptide Fractionation, Mass Spectrometry, and Data Analysis

An amount of 100 µg each of the 0, 16, 24, 48, and 72 hour doxycycline-inducible GATA4 lysates as well as the XEN lysates were prepared for labeling with six-plex TMT reagents according to the manufacturer's instructions (Thermo Fisher, Scientific, Loughborough, Leics., UK, www.lifetechnologies.com). TMT labeling was performed separately for each replicate experiment, and the order of the tags was reversed for one of the replicates. The same time-points for the GATA6-induced time-course were also separately labeled and processed for mass spectrometry.

The pooled TMT-labeled sample was fractionated by off-line high pH reversed-phase liquid chromatography, using an Acquity UPLC system (Waters, Milford, MA, www.waters.com) followed by a second dimension of chromatographic separation using a nanoAcquity UPLC (Waters, Milford, MA, www.waters.com) directly coupled to an LTQ Orbitrap Velos (Thermo Fisher Scientific, Hemel Hempstead, UK, www.thermofisher.com). Proteome Discoverer (version 1.3.0.339, Thermo Fisher Scientific, Hemel Hempstead, UK, www.thermofisher.com) was used for data processing. Spectra were searched against a Mascot search engine (version 2.3.02, Matrix Science, London, UK, www.matrixscience.com/) with a Swiss-Prot mouse database containing 24,473 entries (www.uniprot.org, March 2013) as well as a list of common laboratory contaminants (cRAP repository v1.0, ftp.thegpm.org/fasta/cRAP in March 2013). See Supporting Information Methods for further details of sample preparation and mass spectrometry.

Multivariate Analysis

The multivariate analysis was conducted using the Bioconductor [19] pRoloc package [20, 21] implemented in the R statistical programming language. Support vector machines (SVM) classification and *k*-means clustering analysis were performed on the GATA4-induced dataset. The ClueGO functional annotation plugin for Cytoscape was used to identify functionally related groups of enriched gene ontology (GO) terms and pathway networks associated with the identified SVM and *k*-means clusters [22]. The CLUSTERnGO (CnG) software and source code to carry out model based clustering are available online (http://www.cmpe.boun.edu.tr/content/CnG) and will be described in detail elsewhere. Biological replicates were treated as replicates in CnG when prompted. See Supporting Information Methods for further details on the multivariate analysis.

Results

Doxycycline-Inducible GATA Expression System to Study Proteome-Level Changes

To investigate the transition from an Epi to a XEN-like state, we used recombinase-mediated cassette exchange to develop ESC lines carrying *Gata4* or *Gata6* cDNAs under the control of doxycycline-inducible promoters [18 and (Wamaitha et al., accepted)]. Addition of doxycycline to the culture medium induced efficient transgene expression in the majority of cells (Fig. 1A, 1B). Western blotting and quantitative mass spectrometry showed that expression levels of inducible GATA4 protein were comparable with those observed in embryo-derived XEN cells (Fig. 1C, 1D). Induction of GATA4 expression led to the rapid and simultaneous downregulation of the pluripotency marker NANOG (Fig. 1C, 1E) and upregulation of the PrE marker SOX17 (Fig. 1C). This doxycycline-inducible expression system therefore recapitulates key findings that were previously obtained using alternative GATA expression systems [11, 12].

Within 3 days of continuous doxycycline treatment, the ESCs gradually assumed a phenotype that closely resembled embryo-derived XEN cells (Fig. 1F; Supporting Information Movie). We decided to delineate the changes in the total proteome during the early phase of this fate transition at 0, 16, 24, 48, and 72 hours after doxycycline addition. In parallel, we monitored an embryo-derived XEN cell line, representing the fully differentiated phenotype. Each sample was proteolytically digested and differentially labeled with amine-reactive TMT (Fig. 2A). This quantitative approach allowed all six samples to be simultaneously coanalyzed by liquid chromatography–mass spectrometry (LC-MS). The resulting dataset provides a dynamic account of the global proteomic changes as cells transit from a pluripotent state towards PrE differentiation.

In three independent biological replicate experiments, a total of 3,044 proteins were quantified. Of these, 2,336 proteins (76%) were robustly quantified at all time-points and in all replicate experiments (Fig. 2B; Supporting Information Tables S2, S3). The individual time-points of all three biological replicates were hierarchically clustered and confirmed excellent reproducibility between independent experiments (Fig. 2C). Furthermore, the first 24 hours of the time-course were distinct from the 48 and 72 hour time-points; while the entire time-course clustered separately to the XEN cells. We conclude that this quantitative proteomics approach can reliably track a large proportion of the proteome and capture gross changes during the transition from an ESC- to a XEN-like phenotype.

Induction of GATA4 or GATA6 Expression Induces Similar Global Proteomic Changes

Previous studies have used both *Gata4* and *Gata6* overexpression to induce a XEN-like phenotype in ESCs, and these studies suggested that both proteins were equivalent as judged by

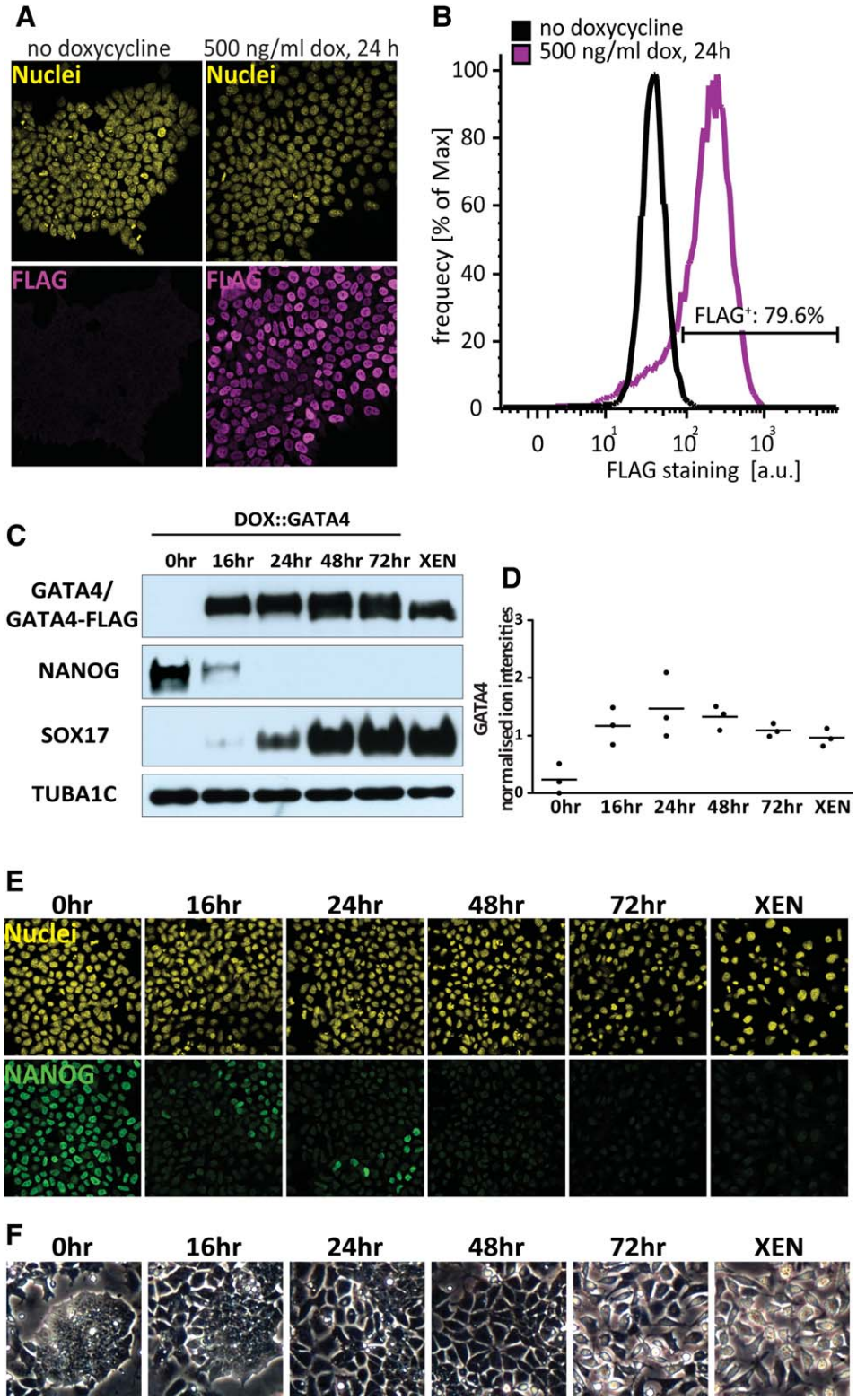

Figure 1. Extra-embryonic differentiation induced by doxycycline-controlled *Gata4* expression. **(A)**: Immunostaining for the FLAG-tag in ESCs carrying a doxycycline-inducible *Gata4-FLAG* transgene 24 hour after addition of doxycycline. The majority of cells express the inducible GATA4-FLAG protein. **(B)**: Flow cytometry of cells treated as in (A) and immunostained for the FLAG tag. 80% of cells express the inducible GATA4-FLAG protein. **(C)**: Western blot analysis of GATA4, NANOG, and SOX17 expression during 3 days of induced GATA4 expression. A strong GATA4 signal, comprising both inducible and endogenous protein is detected within 16 hours of doxycycline-induction, followed by rapid downregulation of the pluripotency marker NANOG and upregulation of SOX17, a marker of primitive endoderm differentiation. α-Tubulin was used as a loading control (TUBA1C) **(D)** GATA4 protein expression measured in the tandem mass tags (TMT)-proteomics dataset (dots indicate normalized reporter ion intensities from independent experiments, bar indicates mean). GATA4 expression reaches a similar level in ESCs carrying the inducible transgene as compared to embryo-derived extra-embryonic endoderm (XEN) cells. Note that both Western blot analysis and TMT ion intensities report both inducible GATA4-FLAG and endogenous GATA4 expression; see Figure 3A for a distinction between induced and endogenous GATA expression. **(E)**: Immunostaining of for NANOG expression in ESCs carrying the doxycycline-inducible *Gata4-FLAG* transgene at different time-points after doxycycline addition. NANOG expression is rapidly and near-simultaneously downregulated in the population. **(F)**: Phase contrast microscopy of cells undergoing the 3-day time-course of GATA4 expression show the morphological changes in cells acquiring a XEN-like phenotype. Abbreviation: XEN, extra-embryonic endoderm.

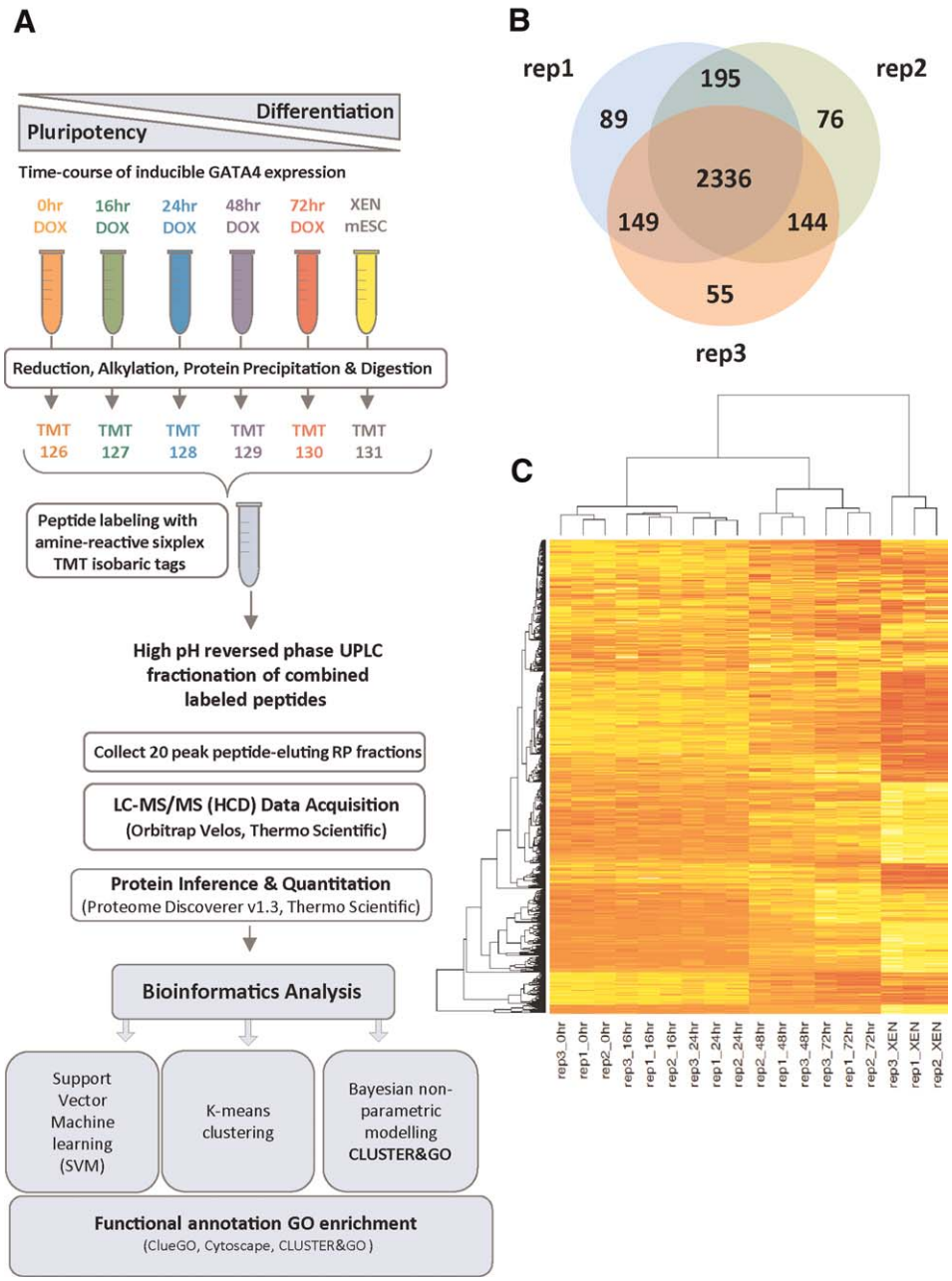

Figure 2. Quantitative proteomics analysis of ESC differentiation following inducible GATA4 expression. **(A)**: Experimental workflow: Five time-points of GATA4 expression (0, 16, 24, 48, and 72 hour) were used to investigate early differentiation events; extra-embryonic endoderm (XEN) cells were used to represent fully differentiated primitive endoderm cells. Cells were harvested and processed for tandem mass tags labeling as described in the Materials and Methods. Downstream bioinformatics analysis was performed with support vector machine classification, *k*-means clustering, model-based clustering, and gene ontology annotation (see Materials and Methods). **(B)**: Venn diagram representing the overlap in proteins that were robustly quantified in the three replicate experiments. A total of 2,336 proteins were quantified in all replicates. **(C)**: Hierarchical clustering demonstrates excellent reproducibility between biological replicate experiments. Abbreviations: DOX, doxycycline; GO, gene ontology; HCD, Higher-energy collisional dissociation; LC-MS, liquid chromatography-mass spectrometry; mESC, mouse embryonic stem cells; SVM, support vector machine; TMT, tandem mass tags; XEN, extra-embryonic endoderm.

marker gene expression and functional tests of the differentiated cells [11, 12]. In the mouse blastocyst, however, *Gata6* and *Gata4* are expressed sequentially [2, 23], and it is possible that subtle functional differences between these factors have been previously missed. To address this question, we subjected ESC lines expressing doxycycline-inducible GATA6 to the same differentiation time-course as the inducible GATA4 lines above. Western blot analysis using FLAG antibodies showed that both inducible GATA4 and GATA6 were expressed with similar dynamics following doxycycline addition, and triggered rapid expression of endogenous GATA factors with comparable efficiency (Fig. 3A), suggesting induced GATA4 and GATA6 have similar functions in inducing PrE-like gene expression.

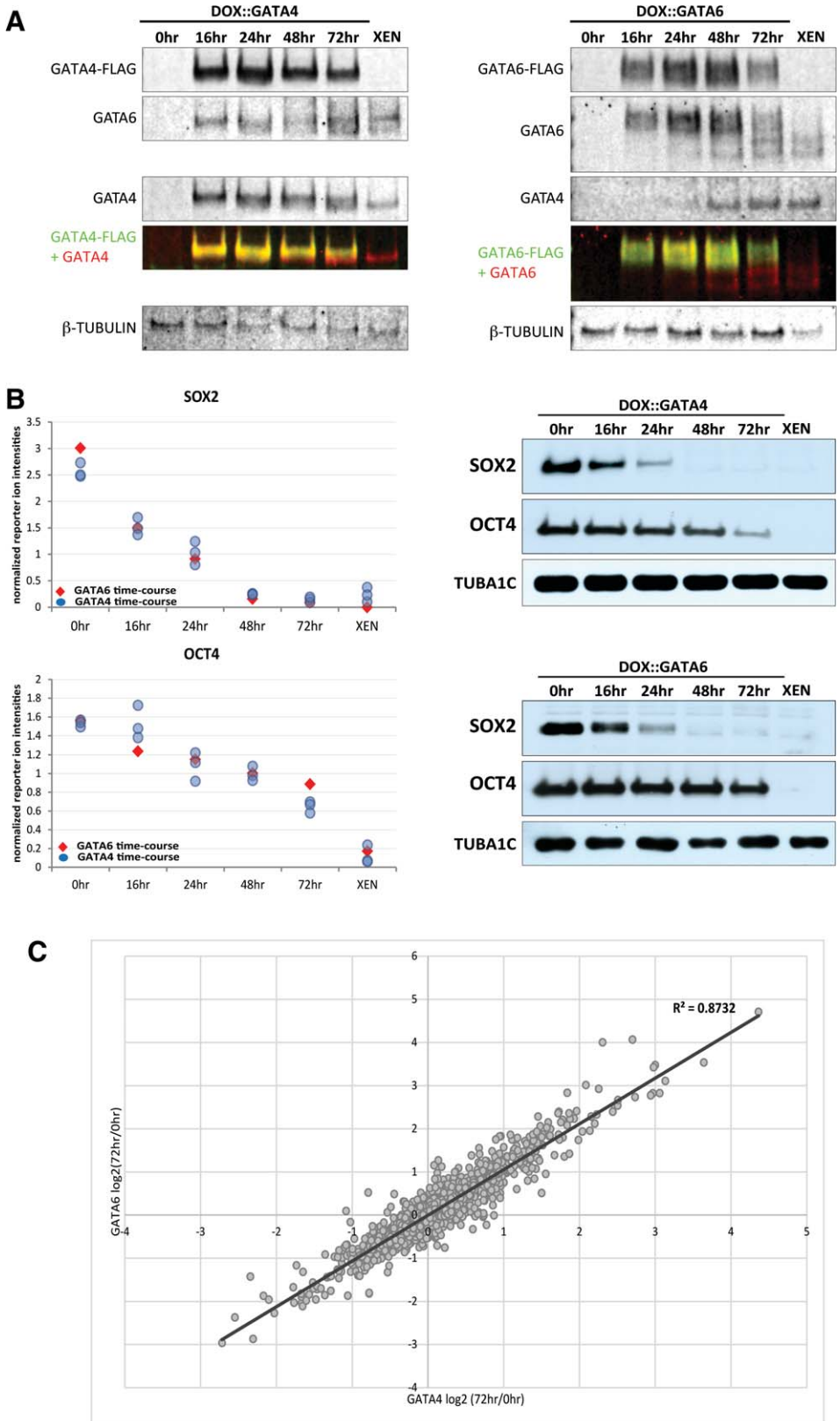

Figure 3. Overexpression of GATA6 induces the same phenotype as GATA4. **(A)**: Comparison of the expression dynamics of inducible and endogenous GATA factors in cell lines carrying an inducible *Gata4-FLAG* (left) or *Gata6-FLAG* (right) transgene. FLAG antibodies (upper panels) specifically detect the inducible protein. Because of its slightly lower molecular weight, simultaneous detection of FLAG-tagged and total GATA protein using secondary antibodies labeled with distinct fluorophores reveals the endogenous protein as a distinct band (red) below the signal from the FLAG-tagged protein in the overlay image (fourth panel from above). Expression of endogenous GATA4 and GATA6 occurs with similar dynamics following doxycycline addition in both cell lines. **(B)**: Left: Scatter plots of the relative protein abundance for SOX2 and OCT4 in each of the three GATA4 replicates (blue circles) and the single GATA6 experiment (red diamond) as determined by mass spectrometry. Right: Western blotting of SOX2 and OCT4 in total lysates obtained from the GATA4 and the GATA6-inducible time-courses. The dynamics of downregulation for these proteins is similar following inducible GATA4 and GATA6 expression. **(C)**: Linear regression analysis for the 1,765 common proteins quantified in the GATA4 triplicate study and the GATA6 experiment. The ratio between the 72-hour time-point and 0 hour time-point (72 hours/0 hour) for the average of the three GATA4 replicates versus the GATA6 experiment was plotted ($r^2 = 0.8574$). The full GATA6 dataset can be found in Supporting Information Table S4. Abbreviations: DOX, doxycycline; XEN, extra-embryonic endoderm.

TMT-labeling and mass spectrometry of the inducible GATA6 line quantified 1,765 proteins that were also identified in all GATA4-inducible replicate experiments (Supporting Information Table S4). Dynamic changes throughout the two time-series were very similar for many proteins of interest (Supporting Information Fig. S1), including OCT4 and SOX2 (Fig. 3B). In fact, the GATA6- and GATA4-induced changes were strongly correlated when comparing the 0 hour and 72 hour time-point for all commonly identified proteins ($r^2 = 0.857$, Fig. 3C). Similar correlation coefficients were found when comparing each of the GATA4 replicates with the GATA6 data (Supporting Information Fig. S2), suggesting that the activity of GATA4 and GATA6 proteins in inducing a XEN-like fate are interchangeable. We therefore decided to focus on the dynamic GATA4-inducible triplicate dataset, and used multivariate bioinformatics tools to assign putative, novel markers of pluripotency and differentiation, as well as to cluster proteins which displayed similar temporal abundance patterns in response to the GATA4 stimulus.

Novel Indicators of Pluripotency and XEN Cell Fate can be Assigned Using a SVM Learning Algorithm

Pluripotency is regulated by a core transcriptional network of three transcription factors, NANOG, SOX2, and OCT4. While we failed to reliably detect NANOG protein in the GATA4-inducible proteomics dataset due to its low abundance, extensive post-translational modification and lack of tryptic sites amenable for mass spectrometry [24], SOX2 was found to rapidly downregulate within the first 24 hours following doxycycline exposure, while OCT4 expression was gradually reduced but still detectable at 72 hours (Fig. 3B; Supporting Information Fig. S3). We reasoned that interrogating the proteomic dataset for proteins with dynamic changes similar to those of the core transcription factors might reveal novel markers of pluripotency and the XEN fate, as well as highlight key molecular pathways underlying the respective states.

The temporal abundance profiles of a selection of well-curated "pluripotency" and "differentiation" marker proteins were used as training data for a SVM learning algorithm implemented in the pRoloc package [20, 21]. Proteins were assigned to clusters defined by these marker proteins based on their similarity to the training data. The SVM scores are a posteriori class probability predictions: high scores are given to proteins which demonstrate similar dynamic abundance profiles to the training set. Using optimized parameters, 88 and 56 proteins were robustly classified to either the "pluripotency" or "differentiation" clusters, respectively (Supporting Information Table S2). Figure 4A depicts the localization of all proteins in the dataset in principle component (PC) space, highlights proteins of the training sets (black circles) and the newly identified members of the pluripotency (blue) and differentiation (red) cluster. The excellent separation of the clusters in PC space suggests that we have identified sets of proteins suited to distinguish the pluripotent state of mouse ESCs from the differentiated XEN-like state.

Proteins in the pluripotency-related cluster downregulate in abundance following the onset of GATA4 expression. 18 proteins within this group were matched to the PluriNetwork, an electronic repository for regulatory networks and molecular mechanisms underlying pluripotency [25] (Supporting Information Table S5). Several other proteins in this cluster have also been linked to pluripotency, such as L1TD1 [26], LYAR [27], DPY30 [28], and TDH [29]. "Heterochromatin" was identified as the most significantly enriched GO term associated with the pluripotency group and many additional chromatin remodeling proteins were identified just below the SVM score threshold (Supporting Information Fig. S4). The temporal abundance pattern for one example from this group, JARID2, was confirmed by Western blotting and immunostaining (Fig. 4B). JARID2 is a key regulator of ESC development and mediates the recruitment of the polycomb repressive complex (PRC2) to target genes [30–33]. Our results reaffirm that the pluripotent state is characterized by a dedicated molecular machinery that maintains a specific chromatin state, which rapidly changes upon PrE differentiation [34].

Inspection of the differentiation cluster suggested that increased deposition of basement membrane and extracellular matrix (ECM) is a hallmark of XEN cell differentiation. While our training set contained a number of ECM components that had previously been implicated in XEN cell differentiation such as COL4A1, COL4A2, LAMA1, and HSPG2 [10], functionally related proteins featured prominently amongst the newly identified components of this cluster, for example, the nidogens NID1 and NID2 and the collagen biosynthetic enzymes PLOD1 and GLT25D1. The differentiation cluster also suggested that components of endocytic protein trafficking pathways are coregulated during XEN differentiation. Our SVM training set included the endocytic receptor cubilin, which is essential for the formation of definitive and visceral endoderm [35], and the adapter protein DAB2 [10, 36]. Several interactors of cubilin and DAB2 were subsequently classified to the differentiation cluster, such as LRP2 (megalin), a major binding partner of cubilin [37, 38] which has been linked to PrE formation [39], and its binding partner LRPAP1, unconventional Myosin-6, which can direct DAB2 trafficking [40], and the LDLR chaperone MESD, which is essential for apical localization of LRP2 in the visceral endoderm [41] (Fig. 4C).

The SVM analysis also highlighted a number of proteins relating to MAPK/ERK signaling. SPROUTY4, for example, inhibits FGF signaling and was identified in the pluripotency-associated group (Fig. 4D). GLIPR2 and FAM129B were both assigned to the differentiation cluster (Fig. 4D) and have recently been identified as direct transcriptional targets of ERK signaling in HK-2 cells and human melanoma cells, respectively [42, 43]. This suggests that cells upregulate FGF/MAPK signaling by shutting down negative feedback components during PrE differentiation, thereby ensuring the continuous FGF/MAPK signaling required for PrE differentiation [44, 45]. Consistent with this idea, we found increased phosphorylation levels of the MAP kinase ERK by immunoblotting (Fig. 4E) along the differentiation time-course, although phosphorylation levels in established XEN cell lines were lower than after 72 hours of doxycycline exposure. Both GLIPR2 and FAM129B have been implicated in endowing cells with increased motility [42, 43], suggesting that they provide a link between the dependence of PrE differentiation on FGF/MAPK signaling [44, 45] and acquisition of a motile phenotype upon differentiation along this lineage.

K-Means Clustering Identifies Global Temporal Abundance Changes

We next performed *k*-means clustering to obtain an overview of all global changes and to identify further clusters of

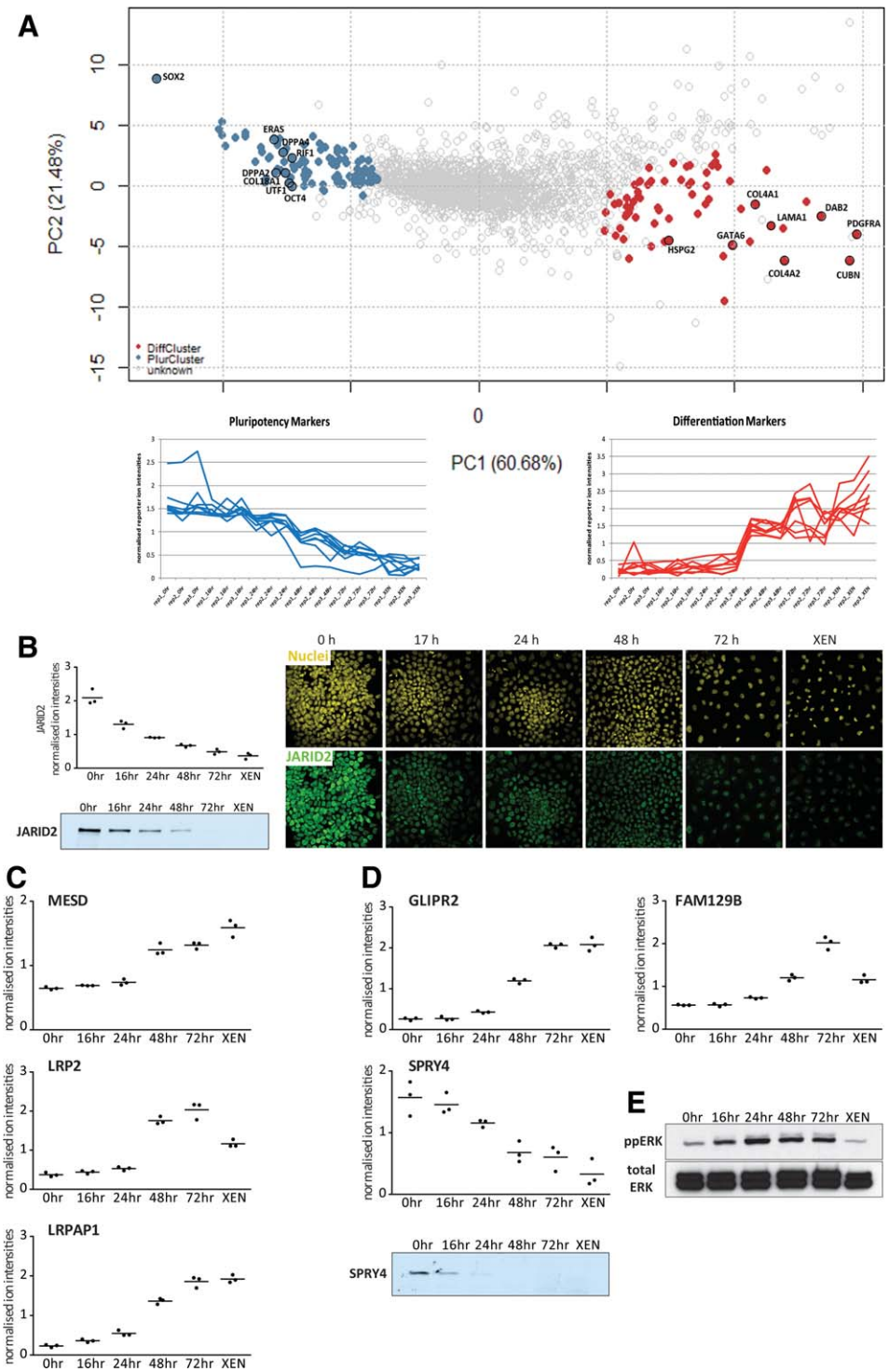

Figure 4. Support vector machine learning algorithm (SVM) identifies novel markers of pluripotency and differentiation. **(A)**: Principle components analysis (PCA) plot of all proteins identified reproducibly upon inducible GATA4 expression. Eight well-curated markers each of pluripotency and differentiation were used as training data for a SVM analysis (circled and labeled in black). Panels to the left and right show the temporal abundance profiles of the 8 pluripotency (blue, left) and differentiation markers (red, right), respectively. Colored spots on the PCA plot indicate proteins assigned to the pluripotency cluster (blue, 88 proteins classified using a threshold at the 95th percentile) and the differentiation cluster (red, 56 proteins classified using a 90th percentile cutoff); see Supporting Information Table S2. Unclassified proteins are shown in grey circles. **(B)**: Tandem mass tags (TMT) reporter ion distribution, Western blot and immunostaining (right), for JARID2, a novel putative marker for the pluripotent state. See Figure 3B for Western blot loading control. **(C)** TMT reporter ion distributions of MESD, LRPAP1, and LRP2, which were classified to the differentiation cluster. **(D)**: Dynamic changes in proteins related to the FGF/MAPK signaling pathway; upper panels: TMT reporter ion distribution for GLIPR2 (left) and FAM129B (right); lower panels: Reporter ion distributions and Western blot for SPRY4. Data points represent the protein abundance in individual experiments, bars indicate mean. See Figure 3B for Western blot loading control. **(E)** Western blot analysis for p42/p44 ERK phosphorylation and total ERK expression levels following induced GATA4-FLAG expression. Abbreviations: ERK, extracellular-signal-regulated kinases; PC, principle component; ppERK, doubly phosphorylated ERK; XEN, extra-embryonic endoderm.

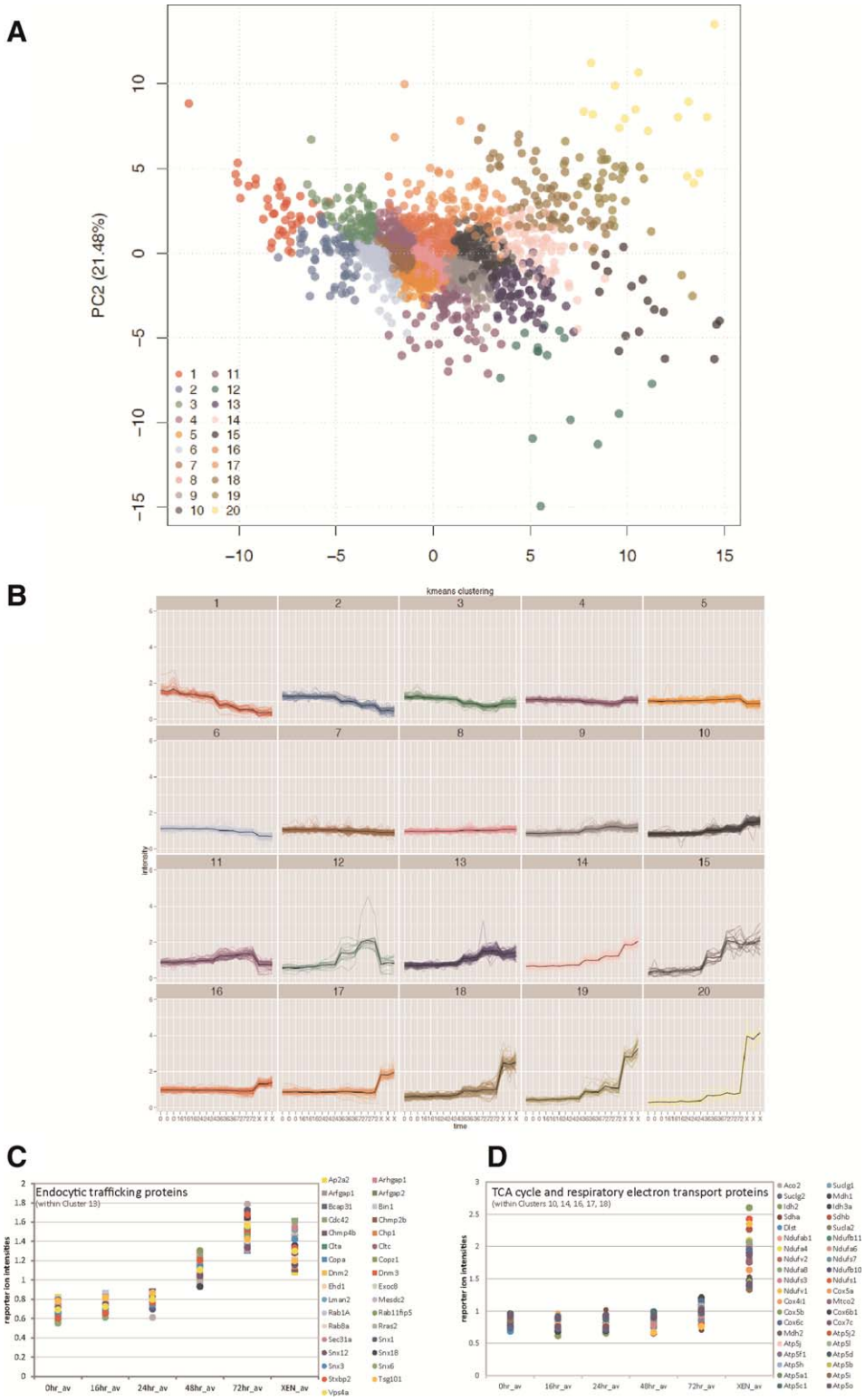

Figure 5. *K*-means clustering reveals distinct clusters in the GATA4-induced dataset. **(A)**: Principle components analysis (PCA) plot of the entire dataset, where each spot represents the tandem mass tags distribution profile for a protein across the time-course. Each spot is colored by membership to one of the 20 clusters identified by *k*-means clustering. **(B)**: Temporal changes in reporter ion intensities for proteins in each cluster. Replicate time-points are represented next to each other. **(C)**: Averaged temporal abundance profiles of selected proteins functionally associated with endocytic trafficking (identified in *k*-means cluster 13). **(D)**: Averaged temporal abundance profiles of selected proteins functionally associated with the citrate acid (TCA) cycle and the respiratory electron transport chain (identified in *k*-means clusters 10, 14, 16, 17, 18). Abbreviations: PC, principle component; TCA, tricarboxylic acid cycle.

Table 1. GO annotation enrichment of *k*-means clusters[a]

K-means cluster No.	No. proteins/ cluster	Enriched term	Term ID	No. proteins/ term	Benjamini-Hochberg corrected *p* value
1	41	Female pronucleus	GO:0001939	3	9.81E-06
2	66	Nuclear chromatin	GO:0000790	7	1.22E-07
3	90	DNA replication	KEGG:03030	11	5.20E-16
4	177	S Phase	REACTOME:5416652	15	8.25E-11
5	196	RNA transport	KEGG:03013	19	7.77E-13
6	236	Ribosome biogenesis in eukaryotes	KEGG:03008	22	1.33E-21
7	392	Gene expression	REACTOME:5416766	176	6.09E-116
8	250	M Phase	REACTOME:5417115	31	3.28E-19
9	160	Membrane Trafficking	REACTOME:5417343	11	1.02E-05
10	136	Mitochondrial inner membrane	GO:0005743	16	2.51E-11
11	66	Cytoskeleton organization	GO:0007010	10	2.38E-05
12	16	Clathrin derived vesicle budding	REACTOME:5417507	2	0.00169532
13	80	Endocytosis	KEGG:04144	17	1.30E-15
14	84	Protein processing in endoplasmic reticulum	KEGG:04141	9	4.13E-07
15	17	Collagen type IV	GO:0005587	2	2.36E-04
16	144	Mitochondrion	GO:0005739	48	7.03E-21
17	66	The citric acid (TCA) cycle and respiratory electron transport	REACTOME:5416705	16	1.27E-16
18	71	Protein processing in endoplasmic reticulum	KEGG:04141	17	6.49E-21
19	33	Protein processing in endoplasmic reticulum	KEGG:04141	5	2.27E-05
20	16	Laminin-10 complex	GO:0043259	2	3.97E-05

[a]ClueGO annotation enrichment analysis was used to identify functional processes associated with each cluster. The ClueGO results for each cluster can be found in Supporting Information Table S6.

proteins with dynamic profiles different from those represented by our SVM training set. Figure 5A depicts the 20 *k*-means clusters in PC space. The temporal profiles for proteins within each cluster shows that they display similar abundance patterns, that is, time of onset and magnitude of response (Fig. 5B, Supporting Information Table S2). We observed good correlation between our supervised (SVM) and unsupervised (*k*-means) clustering methods (Supporting Information Fig. S5).

To obtain a functional insight into the temporal order of events underlying the early phases of fate transition, we performed GO enrichment on all *k*-means clusters (Table 1; Supporting Information Table S6). Expression of proteins in Cluster 2, for example, was initially unaffected by GATA4 induction, but these proteins were subsequently downregulated at 48 hours to a very low level. This cluster was most enriched for the term "nuclear chromatin". Conversely, Cluster 13 was upregulated at 48 hours of differentiation and was highly enriched for "endocytosis" and "membrane trafficking," including proteins such as clathrins, coatamers, sorting nexins, exocyst docking proteins, and endosomal sorting proteins (Fig. 5C). This demonstrates that reorganization of the endomembrane system occurs after a lag time during the transition towards a XEN-like state.

K-means analysis also revealed a number of clusters whose relative abundance was much greater in XEN cells than during the GATA4 time-course (Clusters 16–20). Overrepresented functional terms for these clusters highlight processes that differ between the embryo-derived XEN cells and the GATA4-induced XEN-like cell type, potentially because these processes are further downstream in the differentiation cascade and not yet activated by 72 hour of GATA4 expression. Clusters 18 and 19 were highly enriched for "protein processing in the endoplasmic reticulum (ER)" and further inspection of the dataset revealed that enzymes involved in fatty acid biosynthesis, glycosylation, unfolded protein response and collagen biosynthesis were upregulated in XEN cells (Supporting Information Table S7). GO term analysis for Clusters 16 and 17 revealed enrichment of the terms "tricarboxylic acid (TCA) cycle and respiratory electron transport" and "mitochondrion" (Table 1). Specifically, we found that XEN cells contained a moderate increase in components of the mitochondrial electron respiratory chain and enzymes involved in the TCA cycle (Fig. 5D). This finding supports the observation that differentiated cells have a higher reliance on oxidative phosphorylation than ESCs, and may undergo mitochondrial biogenesis to maintain their increased metabolic demands [46]. Consistent with the fact that ESCs rely predominantly on glycolysis, we find a gradual decrease in expression of the rate-limiting glycolytic enzyme hexokinase-2 (HK2) and SLC2A1 (GLUT1), the major facilitative glucose importer in ESCs, upon differentiation. Many other glycolytic enzymes, however, are unchanged or even upregulated in our dataset (e.g., HK1, PFKL, GPI, PGK1), suggesting there may be more complex adjustments to the metabolic status of these cells.

As expected, many proteins remained unchanged throughout the experiment (e.g., Cluster 7 and 8). We also found two clusters (11 and 12) that were upregulated during the time-course but remained at very low abundance in XEN cells. A longer time-course would be required to distinguish whether these proteins are transiently expressed during the early phase of the transition, or whether they represent effects that specifically result from the long-term exposure of cells to the doxycycline-induced GATA factors.

Additional Biologically Informative Clusters are Revealed Using the Novel Model-Based Algorithm CLUSTERnGO

The *k*-means clustering results suggested there is an almost-continuous distribution of dynamic profiles in our dataset that may not be fully resolved within 20 discrete clusters. To address this, we applied CLUSTERnGO (CnG), a model-based clustering algorithm specifically designed for analyzing

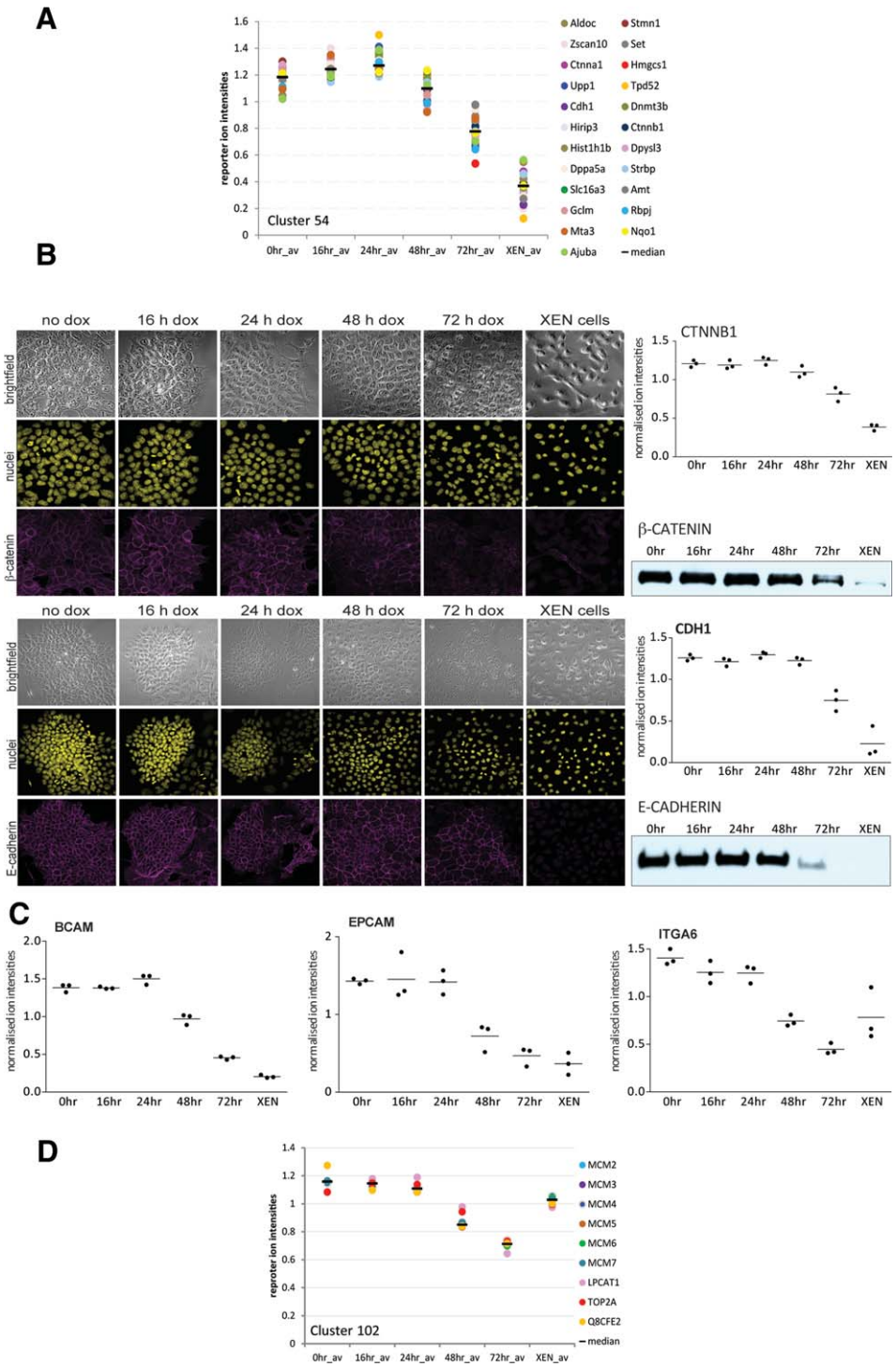

Figure 6. CLUSTERnGO reveals additional biologically relevant clusters. **(A)**: Averaged temporal abundance profiles of proteins in Cluster 54 of the CnG analysis. This cluster contained 23 proteins, including the core adherens junction components α-catenin, β-catenin and cadherin-1 (E-cadherin). **(B)**: Confocal microscopy and Western blotting of β-catenin (CTNNB) and cadherin-1 (CDH1) confirms that they are reduced rapidly in abundance from 48 hours of GATA4 induction. **(C)**: Tandem mass tags reporter ion profiles for BCAM, EPCAM, and ITGA6. Spots represent the protein abundance in three replicate experiments, bars indicate mean. **(D)**: CnG Cluster 102, containing nine proteins, six of which are components of the MCM2-7 complex. Shown are the averaged temporal abundance profiles for the three GATA4 replicate experiments combined. Abbreviations: BCAM, basal cell adhesion molecule; CDH, cadherin-1; CTNNB, β-catenin; EPCAM, epithelial cell adhesion molecule; ITGA6, integrin-α6; XEN, extra-embryonic endoderm.

dynamic time-series datasets that does not place any constraints on the size or number of clusters assigned (see http://www.cmpe.boun.edu.tr/content/CnG for further details and source code). In this way, we hoped to identify subsets of proteins with strongly correlated reporter ion profiles that may have been obscured in the *k*-means approach.

Using optimized parameters, this analysis revealed 206 distinct multiprotein clusters (average of 16 proteins per cluster) with only 30 proteins assigned to single-member clusters (Supporting Information Tables S2, S8). Cluster 54 (Fig. 6A) containing 23 proteins, was initially unchanged in abundance by the differentiation signal, but subsequently was rapidly downregulated at 48 hours. All proteins in this cluster were assigned to the larger *k*-means cluster 2 containing 66 proteins. Cluster 54 includes ZSCAN10, a transcription factor known to play a key role in maintaining pluripotency [47], α-catenin, β-catenin, and cadherin-1 (E-cadherin) which are core components of the adherens junctions, as well as AJUBA, a LIM-domain protein which colocalizes with cadherins and catenins [48, 49]. We confirmed this pattern of temporal abundance changes for β-catenin and cadherin-1 by Western blotting and immunostaining (Fig. 6B). The cell adhesion molecules EPCAM, BCAM, and integrin-α6 (ITGA6) display a similar pattern of expression although they were not assigned to the same cluster (Fig. 6C). Interestingly, the time-point at which we begin to detect downregulation of these junctional proteins (48 hour) coincides with distinct morphological changes such as breakdown of cell-to-cell contact and dispersal of the tightly adherent ESC colonies (Supporting Information Movie). This congruence between the dynamics of phenotypic change and the behavior of proteins within one specific cluster suggests that this clustering approach allows us to infer a temporal order of events during the early stages of differentiation.

CnG Cluster 102 is composed of nine tightly coregulated proteins and contains all six components of the minichromosome maintenance complex (MCM2-7) which is essential for DNA replication initiation, as well as DNA topoisomerase 2-α, an important regulatory enzyme for DNA replication and transcription (Fig. 6D). Seven of the nine proteins in this cluster had been assigned to the *k*-means Cluster 3, which was itself enriched for the GO term "DNA replication." Taken together, clustering by *k*-means and CnG suggests that a group of proteins involved in the control of transcription and maintaining genomic integrity is gradually downregulated as cells differentiate towards extra-embryonic fates, reflecting a slight decline in proliferation over the course of the experiment.

Partitioning the dataset into a larger number of clusters using CnG has identified small cohorts of tightly coregulated proteins, some of which had been overlooked in the *k*-means analysis. The majority of CnG groups, however, were found to be subsets of those previously identified with *k*-means, indicating that the major biological findings from this dataset are not dependent on the clustering method used.

Discussion

Understanding the mechanisms that underlie the functional specialization of stem cells as they differentiate is of paramount importance for the field of stem cell biology and regenerative medicine. In this study, we have made use of the fact that ESCs can be differentiated in an efficient and synchronous manner towards PrE-like lineages through the induced expression of GATA transcription factors [11, 12], and used this system to investigate with high temporal resolution the dynamic changes that occur in the cell's proteome as they alter their fate and function.

Isobaric labeling and mass spectrometry-based proteomic profiling coupled with multivariate data analysis techniques allowed us to capture the dynamic and sequential response of over 2,000 proteins following GATA4 induction. SVM classification prioritized proteins of interest from a large list of candidates by identifying proteins whose expression dynamics are highly correlated to well-curated markers of pluripotency or differentiation. Functional validation experiments will be necessary to further refine this group and to establish their relationship to each other and to the core pluripotency transcription factors.

The high temporal resolution of our dataset allowed us to define the sequences of change between the pluripotent ESC state and extra-embryonic differentiation. Both *k*-means clustering and a novel, model-based clustering approach identified temporally coregulated proteins. For example, we found that one of the first global changes following GATA4 induction is alteration of the chromatin-regulatory network. Later along the differentiation time-course, we detected loss of the expression of the core components of adherens junctions and cell adhesion molecules and a concomitant increase in the deposition of ECM and basement membrane proteins, corresponding with alterations in the morphology of the cells and a coincident increase in cell motility. Together, these observations suggest that our differentiation time-course captures an epithelial–mesenchymal transition (EMT). ESCs form adherent cell sheets in culture, and the PrE of the pre-implantation blastocyst and the visceral endoderm of the post-implantation embryo display distinct epithelial characteristics whereas the parietal endoderm is a dispersed cell population specialized in ECM production. Because embryo-derived XEN cells and XEN-like cells obtained by induced GATA factor expression mostly contribute to the parietal endoderm [12, 50], it appears likely that the EMT observed in our dataset recapitulates events underlying the emergence of the parietal endoderm in the embryo. Importantly, EMTs are a recurrent feature of cell fate conversions during embryonic development, and also play an important role in tumor progression and metastasis. The EMT-like changes detected in our dataset occur simultaneously with an upregulation in membrane trafficking and defined metabolic alterations. It will be interesting to investigate whether this is a general theme in different EMTs during differentiation and disease.

Global cellular changes in the transition from an ESC to a XEN-like state have recently been investigated at the transcriptomic level using inducible *Sox17* expression [50]. This study revealed transcript-level alterations in several functional processes that were also implicated in our GATA4-inducible proteomics dataset. Inducible expression of *Sox17*, however, appears to trigger extra-embryonic differentiation with a different dynamic signature to that observed in the embryo and upon inducible GATA expression, making it difficult to directly compare the results of this study with ours.

While we have delineated the major proteomic changes that occur as cells differentiate and restrict their lineage potential, a recent publication examined the temporal proteomic dynamics of the reverse process, that is, when cells acquire pluripotency through the reprogramming of mouse embryonic fibroblasts (MEFs) to induced pluripotent stem cells (iPSCs) [51]. Surprisingly, the functional processes that underlie the intermediate phase of this transition are very similar

to those identified in our study. Hansson *et al.*, demonstrate that cellular reprogramming involves changes in the stoichiometry of electron transport chain complexes, a switch from mitochondrial oxidation in MEFs to glycolysis in iPSCs, chromatin remodeling, altered vesicle-mediated transport, protein processing in the ER, and an EMT-like process involving changes in ECM and cell adhesion molecules.

Conclusion

In summary, the temporal dissection of our proteomic dataset has revealed sequential changes in the differentiating proteome that may represent crucial hallmarks of the transition from a pluripotent to an extra-embryonic, differentiated state. This dataset provides a valuable insight into lineage determination during early embryogenesis and will serve as an important resource for the stem cell community.

Data Submission

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD001901 and 10.6019/PXD001901 [52, 53].

Acknowledgments

We thank Professor Steve Oliver and Dr. A.K. Hadjantonakis for helpful discussions and advice. This work was supported by the European Union 7th Framework Program (PRIME-XS project Grant 262067 [to K.S.L., L.G, and C.M.M]), the Biotechnology and Biological Sciences Research Council (BBSRC Grant BB/L002817/1 [to K.S.L and L.G.]), a HFSP Grant (RGP0029/2010) and a European Research Council (ERC) Advanced Investigator Grant (to A.M.A.), an EMBO long-term fellowship and a Marie Curie IEF Grant (to C.S.), and by the Medical Research Council (MRC, UK, MC_UP_1202/9) and the March of Dimes Foundation (FY11-436) (to L.T.Y.C. and K.K.N.).

Author Contributions

C.M.M.: designed and planned the experiments, performed the proteomics sample preparation, mass spectrometry, data analysis and Western blotting, wrote the manuscript and approved the final manuscript; C.S.: designed and planned the experiments, performed the cell culture, transfections, FACs analysis and immunostaining, wrote the manuscript and approved the final manuscript; L.G.: performed the multivariate data analysis and submission to the ProteomeXchange repository, approved the final manuscript; D.D. and I.B.F.: performed the CLUSTERnGO analysis, approved the final manuscript; A.C. assisted with data analysis, approved the final manuscript; M.J.D.: assisted with mass spectrometry, approved the final manuscript; L.T.Y.C. and K.K.N.: generated the GATA4 and GATA6 cell lines and performed initial characterization of the cell culture system, approved the final manuscript; A.M.A., and K.S.L.: designed and planned the experiments, wrote the manuscript and approved the final manuscript. C.M.M. and C.S. contributed equally to this article.

Disclosure of Potential Conflicts of Interest

The authors indicate no potential conflicts of interest.

References

1 Chazaud C, Yamanaka Y, Pawson T et al. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell 2006;10:615–624.

2 Plusa B, Piliszek A, Frankenberg S et al. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development 2008;135:3081–3091.

3 Niakan KK, Ji H, Maehr R et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes Dev 2010;24:312–326.

4 Rossant J, Tam PP. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development 2009;136:701–713.

5 Gasperowicz M, Natale DR. Establishing three blastocyst lineages--then what? Biol Reprod 2011;84:621–630.

6 Kunath T, Arnaud D, Uy GD et al. Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development 2005;132:1649–1661.

7 Brown K, Legros S, Artus J et al. A comparative analysis of extra-embryonic endoderm cell lines. PloS One 2010;5:e12016.

8 Beddington RS, Robertson EJ. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development 1989;105:733–737.

9 Morgani SM, Canham MA, Nichols J et al. Totipotent embryonic stem cells arise in ground-state culture conditions. Cell Rep 2013;3:1945–1957.

10 Cho LT, Wamaitha SE, Tsai IJ et al. Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. Development 2012;139:2866–2877.

11 Fujikura J, Yamato E, Yonemura S et al. Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev 2002;16:784–789.

12 Shimosato D, Shiki M, Niwa H. Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. BMC Dev Biol 2007;7:80.

13 Lu R, Markowetz F, Unwin RD et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature 2009;462:358–362.

14 Hansson J, Krijgsveld J. Proteomic analysis of cell fate decision. Curr Opin Genet Dev 2013;23:540–547.

15 Rugg-Gunn PJ, Cox BJ, Lanner F et al. Cell-surface proteomics identifies lineage-specific markers of embryo-derived stem cells. Dev Cell 2012;22:887–901.

16 Frohlich T, Kosters M, Graf A et al. iTRAQ proteome analysis reflects a progressed differentiation state of epiblast derived versus inner cell mass derived murine embryonic stem cells. J Proteom 2013;90:38–51.

17 Rigbolt KT, Prokhorova TA, Akimov V et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal 2011;4:rs3.

18 Beard C, Hochedlinger K, Plath K et al. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis 2006;44:23–28.

19 Gentleman RC, Carey VJ, Bates DM et al. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

20 Gatto L, Breckels LM, Wieczorek S et al. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. Bioinformatics 2014;30:1322–1324.
21 Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 2012;28: 288–289.
22 Bindea G, Mlecnik B, Hackl H et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091–1093.
23 Artus J, Piliszek A, Hadjantonakis AK. The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. Dev Biol 2011;350:393–404.
24 Brumbaugh J, Russell JD, Yu P et al. NANOG is multiply phosphorylated and directly modified by ERK2 and CDK1 in vitro. Stem Cell Rep 2014;2:18–25.
25 Som A, Harder C, Greber B et al. The PluriNetWork: An electronic representation of the network underlying pluripotency in mouse, and its applications. PloS One 2010; 5:e15165.
26 Wong RC, Ibrahim A, Fong H et al. L1TD1 is a marker for undifferentiated human embryonic stem cells. PloS One 2011;6:e19355.
27 Li H, Wang B, Yang A et al. Ly-1 antibody reactive clone is an important nucleolar protein for control of self-renewal and differentiation in embryonic stem cells. Stem Cells 2009;27:1244–1254.
28 Jiang H, Shukla A, Wang X et al. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell 2011;144:513–525.
29 Wang J, Alexander P, Wu L et al. Dependence of mouse embryonic stem cells on threonine catabolism. Sci 2009;325:435–439.
30 Peng JC, Valouev A, Swigut T et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 2009;139:1290–1302.
31 Shen X, Kim W, Fujiwara Y et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 2009;139:1303–1314.
32 Landeira D, Sauer S, Poot R et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol 2010;12:618–624.
33 Pasini D, Cloos PA, Walfridsson J et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 2010;464:306–310.
34 Santos J, Pereira CF, Di-Gregorio A et al. Differences in the epigenetic and reprogramming properties of pluripotent and extra-embryonic stem cells implicate chromatin remodelling as an important early event in the developing mouse embryo. Epigenet Chromat 2010;3:1.
35 Smith BT, Mussell JC, Fleming PA et al. Targeted disruption of cubilin reveals essential developmental roles in the structure and function of endoderm and in somite formation. BMC Dev Biol 2006;6:30.
36 Yang DH, Cai KQ, Roland IH et al. Disabled-2 is an epithelial surface positioning gene. J Biol Chem 2007;282:13114–13122.
37 Cases O, Perea-Gomez A, Aguiar DP et al. Cubilin, a high affinity receptor for fibroblast growth factor 8, is required for cell survival in the developing vertebrate head. J Biol Chem 2013;288:16655–16670.
38 Fisher CE, Howie SE. The role of megalin (LRP-2/Gp330) during development. Dev Biol 2006;296:279–297.
39 Gerbe F, Cox B, Rossant J et al. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. Dev Biol 2008;313:594–602.
40 Maurer ME, Cooper JA. Endocytosis of megalin by visceral endoderm cells requires the Dab2 adaptor protein. J Cell Sci 2005; 118:5345–5355.
41 Lighthouse JK, Zhang L, Hsieh JC et al. MESD is essential for apical localization of megalin/LRP2 in the visceral endoderm. Dev Dyn 2011;240:577–588.
42 Huang S, Liu F, Niu Q et al. GLIPR-2 overexpression in HK-2 cells promotes cell EMT and migration through ERK1/2 activation. PloS One 2013;8:e58574.
43 Old WM, Shabb JB, Houel S et al. Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. Mol Cell 2009;34:115–131.
44 Yamanaka Y, Lanner F, Rossant J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development 2010;137:715–724.
45 Nichols J, Silva J, Roode M et al. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 2009;136:3215–3222.
46 Varum S, Rodrigues AS, Moura MB et al. Energy metabolism in human pluripotent stem cells and their differentiated counterparts. PloS One 2011;6:e20914.
47 Yu HB, Kunarso G, Hong FH et al. Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. J Biol Chem 2009;284:31327–31335.
48 Nola S, Daigaku R, Smolarczyk K et al. Ajuba is required for Rac activation and maintenance of E-cadherin adhesion. J Cell Biol 2011;195:855–871.
49 Marie H, Pratt SJ, Betson M et al. The LIM protein Ajuba is recruited to cadherin-dependent cell junctions through an association with alpha-catenin. J Biol Chem 2003; 278:1220–1228.
50 McDonald AC, Biechele S, Rossant J et al. Sox17-Mediated XEN Cell Conversion Identifies Dynamic Networks Controlling Cell-Fate Decisions in Embryo-Derived Stem Cells. Cell Rep 2014;9:780–793.
51 Hansson J, Rafiee MR, Reiland S et al. Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. Cell Rep 2012;2:1579–1592.
52 Vizcaino JA, Cote RG, Csordas A et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: Status in 2013. Nucl Acid Res 2013;41:D1063–D1069.
53 Vizcaíno JA, Deutsch EW, Wang R, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 2014;32:223–226.


See www.StemCells.com for supporting information available online.